

Gewichtungen zur Anpassung an Statusvariablen: eine Untersuchung am Allbus 1986

Rothe, Günter

Veröffentlichungsversion / Published Version
Forschungsbericht / research report

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Rothe, G. (1989). *Gewichtungen zur Anpassung an Statusvariablen: eine Untersuchung am Allbus 1986*. (ZUMA-Arbeitsbericht, 1989/21). Mannheim: Zentrum für Umfragen, Methoden und Analysen -ZUMA-. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-67041>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

**Gewichtungen zur Anpassung
an Statusvariablen -
Eine Untersuchung am ALLBUS 1986**

Günter Rothe

**ZUMA-Arbeitsbericht Nr. 89/21
November 1989**

Seit Juli 1983 sind die ZUMA-Arbeitsberichte in zwei Reihen aufgeteilt:

Die ZUMA-Arbeitsberichte (neue Folge) haben eine hausinterne Begutachtung durchlaufen und werden vom Geschäftsführenden Direktor zusammen mit den übrigen Wissenschaftlichen Leitern herausgegeben. Die Berichte dieser Reihe sind zur allgemeinen Weitergabe nach außen bestimmt.

Die ZUMA-Technischen Berichte dienen zur hausinternen Kommunikation bzw. zur Unterrichtung externer Kooperationspartner. Sie sind nicht zur allgemeinen Weitergabe bestimmt.

Gewichtungen zur Anpassung an Statusvariablen -
Eine Untersuchung am ALLBUS 1986.

G. Rothe

1. Problemstellung. Ein wesentliches Ziel von Umfragen auf Stichprobenbasis ist es, aufgrund der hierdurch gewonnenen Daten Rückschlüsse auf die Population ziehen zu können, aus der die jeweilige Stichprobe gewonnen wurde.

In der Regel stellen jedoch die aus Umfragen resultierenden Datensätze kein "repräsentatives" Bild der Grundgesamtheit dar - zumindest nicht in dem strengen Sinn, daß die Merkmalsverteilungen und damit alle relativen Häufigkeiten oder Populationsmittelwerte von Variablen, die in der Stichprobe erhoben wurden, identisch sind mit den entsprechenden Größen in der Gesamtpopulation. Dies gilt bereits nicht bei klassischen einfachen Zufallsstichproben, bei denen derartige Anteile bzw. Mittelwerte für die entsprechenden Populationsparameter zwar erwartungstreue Schätzer sind, jedoch in Abhängigkeit vom Umfang der Stichprobe mehr oder weniger um den wahren Populationswert streuen. Bei komplexeren Stichprobenplänen jedoch, wie sie derzeit in der Marktforschung in der Regel zur Anwendung gelangen, ist die Wahrscheinlichkeit, in die Stichprobe zu gelangen, nicht - wie bei einer einfachen Zufallsstichprobe - für jedes Individuum der Grundgesamtheit gleich groß. Eine einfache Hochrechnung von Populationsanteilen oder Mittelwerten würde in einem solchen Fall in der Regel zu verzerrten Ergebnissen führen.

In diesen Fällen ist eine gewichtete Hochrechnung erforderlich. Dem aus der Umfrage resultierenden Datensatz werden hierzu in der Regel eine oder gar mehrere sogenannte Gewichtsvariablen zugespielt, die angeblich exaktere Hochrechnungen der Populationsparameter ermöglichen sollen.

Bei diesen Gewichtsvariablen sind insbesondere zwei Arten zu unterscheiden:

- Die "Transformation". Dieses Gewicht ist proportional zum Kehrwert der jeweiligen Auswahlwahrscheinlichkeit eines Falles im Sinne des vorgegebenen Stichprobenplanes. Es ist theoretisch begründet durch das sogenannte Horvitz-Thompson-Verfahren (vgl. hierzu auch Rothe und Wiedenbeck (1988)).
- Das Redressement. Dieses Gewicht stellt eine Anpassung an externe Häufigkeitsverteilungen dar: Oft ist die Verteilung einer Variablen oder Variablenkombination des Umfragedatensatzes aus einer externen, in der Regel verlässlicheren Quelle verfügbar (z.B. Mikrozensus, Volkszählung). Redressementgewichte, deren Konstruktionsprinzip in einem späteren Abschnitt beschrieben wird, sind so beschaffen, daß bei ihrer Verwendung die Diskrepanz zwischen Stichproben- und Mikrozensus-Verteilung verschwindet oder zumindest vernachlässigbar klein wird. Ist der Stichprobenplan korrekt realisiert, stellt das Redressement formal eine nachträgliche Schichtung dar, die in der Regel die Qualität der Hochrechnungen sogar wirklich verbessert.

Tatsächlich jedoch ist es i.a. unmöglich, einen Stichprobenplan korrekt durchzuführen, da seine Realisierung aus den unterschiedlichsten Gründen beeinträchtigt wird, so z.B. durch:

- Ausfall einiger Interviewer und damit einiger Regionen,
- unkorrekte Ermittlung der Kontaktperson durch den Interviewer,
- Nichterreichbarkeit der Kontaktperson,
- die Verweigerung des Interviews durch die Kontaktperson oder den Haushalt und nicht zuletzt durch
- die (aus Kostengründen) erforderliche Beschränkung der Feldzeit, die in der Regel nur eine geringe, fest vorgegebene Zahl von (meist 3 oder 4) Kontaktversuchen zuläßt.

Damit ist aber automatisch die Verwendung des Transformationsgewichtes, das sich ja an einer exakten Realisierung der Stichprobe orientiert, inkorrekt, da die angesprochenen Ausfallmechanismen in den Stichprobenplan nicht mit einbezogen sind. Im Falle von Variablen, bei denen ein Vergleich mit verlässlicheren Zahlen möglich ist, zeigt sich auch, daß die Transformation den Schätzfehler oft eher noch verschlimmert.

Durch die Durchführung eines Redressements dagegen wird nach wie vor bewirkt, daß zumindest die Verteilung der bei der Adaption berücksichtigten

Variablen mit der gewichteten Hochrechnung nahezu übereinstimmt. Dies wird von den Umfrageinstituten oft als Argument dafür genutzt, daß auch bei anderen Variablen, die nicht in die Adaption eingehen, die Redressementgewichtung zu einer Verbesserung der Qualität beiträgt. Diese Behauptung ist jedoch falsch. Der Einfluß solcher Gewichtungen auf andere Variablen hängt von einer Vielzahl nicht kontrollierbarer Faktoren ab; eine allgemeine Aussage über Gewichtungseffekte ist nicht möglich.

Besonders auffällig sind die Stichprobendefekte unter anderem bei der Alters- und Geschlechtsstruktur der Bevölkerung; es ist z.B. bekannt, daß ältere (alleinstehende) Frauen aufgrund geringer Kooperationsbereitschaft nur sehr schwer erreichbar sind und damit in der Regel in der Stichprobe unterrepräsentiert sind. Die Alters- und Geschlechtsstruktur wird daher beim Redressement der Marktforschungsinstitute stets berücksichtigt. Darüberhinaus jedoch besteht ferner eine deutliche Diskrepanz zwischen der Verteilung von Statusvariablen im Mikrozensus und bei Umfragen; Bevölkerungsgruppen mit niedrigem Schulabschluß und Arbeiter sind in der Regel stark unterrepräsentiert ("Statusverzerrung", oft fälschlicherweise auch als "Mittelschichtsbias" bezeichnet). Dieser Verteilungsdefekt verschwindet in der Regel auch dann nicht, wenn zur Gewichtung die Redressementgewichte der Marktforschungsinstitute herangezogen werden.

Ein typisches Beispiel für eine Umfrage, bei der die angesprochenen Probleme auftreten, stellt der ALLBUS dar (Allgemeine Bevölkerungsumfrage der Sozialwissenschaften), der als gemeinsames Vorhaben von ZUMA und den Zentralarchiv für empirische Sozialforschung (ZA) alle zwei Jahre mit einem teils konstanten, teil variablen Fragenprogramm durchgeführt wird. Der ALLBUS wird in der Regel unter Verwendung des Stichprobenplanes der "Arbeitsgemeinschaft deutscher Marktforschungsinstitute" (AdM) realisiert. Er enthält einen umfangreichen Block soziodemographischer Variablen, der mit Daten des Mikrozensus abgeglichen werden kann, um einen Einblick in die Qualität der Realisierung des Stichprobendesigns zu gewinnen. Das AdM-Design ist ein dreistufiger Stichprobenplan, der in seinen Kernzügen etwa bei Kirschner (1981) im Detail beschrieben ist.

Neben den von den Marktforschungsinstituten konstruierten Redressementgewichten für personen- und haushaltsbezogene Variablen findet sich in den ALLBUS-Datensätzen ferner ein sogenanntes ZUMA-Gewicht. Dieses Gewicht, das

ebenfalls bei Kirschner (1981) beschrieben ist, soll einen speziellen Designdefekt korrigieren und verwendet hierzu eine Modellannahme, von der sich allerdings inzwischen gezeigt hat, daß sie nicht mehr aufrechtzuerhalten ist. In Abschnitt 2 werden wir das AdM-Design und das ZUMA-Gewicht etwas eingehender ansprechen.

Bei allen Gewichtungen, die in den ALLBUS-Datensätzen zu finden sind, bleibt tatsächlich die oben angesprochene Statusverzerrung unverändert. Diese Tatsache wird von vielen Nutzern als unbefriedigend empfunden, da ganz offenbar im ALLBUS-Datensatz designdefektbedingte Verzerrungen auftreten. Diese Defekte lassen sich feststellen, solange ein Vergleich mit externen Datenquellen möglich ist. Sie sind damit natürlich auch vorhanden bei Verteilungen von Variablen, die sich selbst nicht extern validieren lassen.

Aus diesem Grund wurde im Rahmen eines ZUMA-Grundlagenforschungsprojekts die Frage aufgeworfen, ob grundsätzlich die Möglichkeit besteht, bei einer Einbeziehung der Bildungsvariablen in die Redressementprozedur derartige Defekte zu heilen oder ihr Ausmaß zu mindern. Daraufhin wurde eine Untersuchung anhand der Daten des ALLBUS 1986 durchgeführt, über die wir im vorliegenden Manuskript berichten. Anhand von Variablen, die nicht in die Redressementprozedur einbezogen werden, die aber trotzdem noch extern (durch den Mikrozensus) validierbar sind, kann die Auswirkung der unterschiedlichen Redressementprozeduren kontrolliert werden.

An der für diese Untersuchung erforderliche Datenbeschaffung, -aufbereitung und -auswertung waren neben dem Autor Peter Hartmann, Achim Koch und Michael Wiedenbeck von ZUMA wesentlich beteiligt.

2. ALLBUS und Gewichtung. Die bisher durchgeführten ALLBUS-Erhebungen erfolgten stets unter Verwendung des AdM-Designs; einige Abweichungen sind auf verschärfte Anforderungen durch die Auftraggeber sowie die unterschiedliche Handhabung des Designs durch die beauftragten Marktforschungsinstitute zurückzuführen. Vereinfachend kann davon ausgegangen werden, daß zunächst eine Stichprobe von Stimmbezirken erhoben wird, wobei die Auswahlwahrscheinlichkeit für einen einzelnen Stimmbezirk proportional zu der Anzahl der Wahlberechtigten dieses Stimmbezirkes ist (dem sogenannten Bedeutungsgewicht, das konkret allerdings aus unterschiedlichen Datenquellen hochgerechnet wird).

Innerhalb jeden Stimmbezirkes dieser Stichprobe soll nun nach einem Zufallsprinzip eine feste Zahl von Haushaltsadressen angegangen werden. Dieses Prinzip wird mit dem Begriff "Random Route" umschrieben, "random" (also zufällig) ist hierbei allerdings nur die Startadresse, danach steht praktisch die Gesamtheit aller anzugehenden Haushalte des Stimmbezirks durch eine (möglichst) präzise Beschreibung des Weges fest. Durch diese Konzeption wird im Idealfall erreicht, daß jeder Haushalt in der Bundesrepublik die gleiche Chance hat, in die Stichprobe zu gelangen. Innerhalb des Haushalts wird nach dem sogenannten "Schwedenschlüssel" eine Person der Grundgesamtheit zufällig ausgewählt und interviewt. Dies bedeutet, daß sich die Auswahlchancen für eine einzelne Person verringern, wenn die Zahl der zur Grundgesamtheit gehörenden Haushaltsmitglieder ("reduzierte Haushaltsgröße") groß ist; konkret ist die Auswahlwahrscheinlichkeit umgekehrt proportional zu dieser reduzierten Haushaltsgröße. Das oben angesprochene Transformationsgewicht entspricht bei diesem Stichprobenplan somit bis auf einen konstanten Faktor genau der jeweiligen reduzierten Haushaltsgröße. Dieser Faktor ist für eine gewichtete Hochrechnung allerdings unwesentlich, da für die Hochrechnung von Populationsanteilen oder Mittelwerten nur die relative Größe der Gewichte zueinander von Belang ist. Einer allgemein anerkannten Konvention folgend werden sämtliche Gewichte in der Regel so normiert, daß ihre Summe gerade der Gesamtzahl der Fälle in der Stichprobe entspricht.

Das im ersten Abschnitt bereits angesprochene ZUMA-Gewicht trägt der Tatsache Rechnung, daß die Zahl der Haushalte, die pro Stimmbezirk angegangen wurden, nicht - wie im AdM-Design vorgesehen - für alle Stimmbezirke identisch ist. Es wird zunächst eine Adaption konstruiert, die genau dann eine korrekte Transformation darstellen würde, wenn die Zahl der tatsächlich erreichten Haushalte mit der im jeweiligen Stimmbezirk durch den Stichprobenplan tatsächlich vorgegebenen Zahl identisch wäre. Anschließend wird auch eine Adaption an Alter und Geschlecht durchgeführt.

Tatsächlich jedoch ist die Zahl der wirklich erreichten Haushalte nicht von vornherein vorgegeben; sie ist durch die Feldzeit bestimmt und damit zufällig: Durch das Random-Route ist eine feste Zahl von Kontakthaushalten vorgegeben; die Gesamtzahl ist jedoch so überhöht, daß das angestrebte Stichproben-Netto (also die Zahl der zu realisierenden Interviews) bei der in der Feldzeit zu erwartenden Ausschöpfung tatsächlich erreichbar ist. Ist der angestrebte Umfang von Interviews erreicht, wird die Erhebung abgebrochen.

Damit sind in der Regel einige Stimmbezirke voll ausgeschöpft, andere dagegen weisen nur wenige realisierte Interviews (oder auch gar keins) auf. Dies läßt vermuten, daß in jedem Stimmbezirk, der nicht vollständig ausgeschöpft ist, eher schwer erreichbare Haushalte vertreten sind, dagegen dennoch vorwiegend die leichter erreichbaren bzw. kooperationswilligen Haushalte in der Stichprobe vertreten und damit dort überrepräsentiert sind. Diese erhalten dann ihrerseits durch die ZUMA-Gewichtung nochmals eine weitere höhere Repräsentation. Es ist völlig unklar, ob hierdurch ggf. weitere Fehler bei der Hochrechnung bewirkt werden.

3. Redressementprozedur. Bevor wir die verschiedenen in dieser Arbeit untersuchten Redressementprozeduren beschreiben, sind zunächst einige allgemeine Bemerkungen über die verwendete Art von Adaption an externe Daten erforderlich. Als Ausgangspunkt dient stets eine Datei, die in Matrixform vorliegt und bei der jede Zeile einem Fall und jede Spalte einer Variablen entspricht. Jede dieser Variablen besitzt eine endliche Menge von Ausprägungen. Jedem Fall (Zeile) und jeder Variablen (Spalte) entspricht genau eine Ausprägung. Fassen wir mehrere Variablen zu einer Variablengruppe oder Variablenkombination zusammen, so besitzt jeder einzelne Fall hinsichtlich dieser Gruppe eine endliche Zahl von Ausprägungskombinationen, d.h. von Tupeln von Ausprägungen.

Bei einer Auswahl von Variablen ist jede Ausprägungskombination denkbar, die im vorliegenden Datensatz in der Regel mehrmals auftritt. Die jeweiligen absoluten Häufigkeiten für eine Merkmalskombination lassen sich somit den einzelnen Kombinationen zuordnen. Das auf diese Weise gewonnene Häufigkeitstabelleau des Datensatzes werden wir im folgenden als Ist-Tafel bezeichnen. Unter einer Tafel allgemein verstehen wir jede Zuordnung der Gesamtheit aller Ausprägungskombinationen als Definitionsbereich mit reellen Zahlen als Wertebereich; so etwa auch die relativen oder gewichteten Häufigkeiten für die jeweiligen Ausprägungskombinationen. Jede mit einer Zahl zu besetzende Stelle der Tafel bezeichnen wir als Zelle.

Die externen Daten liegen in der Regel nicht fallweise vor, sondern nur in Form einer Tafel absoluter bzw. relativer Häufigkeiten. Die naheliegendste Gewichtungsprozedur besteht nun schlicht in einem sogenannten "Soll-Durch-Ist (SdI)": In jeder Zelle dividiere man die relative Häufigkeit der exter-

nen Tafel (Soll-Tafel) durch die jeweilige relative Häufigkeit in der Ist-Tafel. Alle Fälle der Ist-Datei, die zu dieser Zelle gehören, erhalten dann das auf diese Art und Weise entstandene Gewicht.

Vor der Untersuchung von Eigenschaften dieses Verfahrens sind zunächst einige technische Details zu diskutieren: Formal ist es zunächst unproblematisch, wenn einige Ist-Zellen durch eine 0 besetzt sind. Hier würde die Soll-durch-Ist-Definition zwar eine Division durch 0 implizieren, da die Ist-Datei jedoch keinen Fall in dieser Zelle enthält, gibt es mit der Verteilung der Gewichte in diesem Punkt keinerlei Probleme. Technisch ebenfalls unproblematisch ist der Fall, daß die Soll-Zelle mit dem Wert 0, die Ist-Zelle dagegen mit einem positiven Wert besetzt ist. Das SdI-Gewicht würde derartige Fälle quasi aus dem Ist-Datensatz eliminieren. Handelt es sich bei der Soll-Tafel tatsächlich um die exakte Verteilung, so kann dieser Fall natürlich grundsätzlich nicht auftreten, in der hier vorliegenden Situation, daß die Soll-Tafel aus den Daten des Mikrozensus gewonnen wird, ist es zunächst nicht prinzipiell auszuschließen - auch der Mikrozensus ist ja nur eine Stichprobe. In den in dieser Arbeit betrachteten Zellen tritt dieses Problem jedoch nie auf.

Sind nun sämtliche Ist-Zellen positiv besetzt, so besitzt die SdI-gewichtete Ist-Tafel folgende Eigenschaften:

- Die Gesamtsumme aller Zellbesetzungszahlen ist identisch mit der Fallzahl des Ist-Datensatzes
- Das Verhältnis der (nicht notwendig ganzzahligen) Zellbesetzungszahlen zur Gesamtsumme der Tafel entspricht exakt den relativen Häufigkeiten der Soll-Tafel.

Sind dagegen einige Ist-Zellen mit 0 besetzt, so gilt dies selbstverständlich ebenfalls für die entsprechenden Zellen der gewichteten Ist-Tafel. Die Gesamtsumme des Tableaus reduziert sich um einen Anteil, der genau dem Anteil derjenigen Soll-Zellen entspricht, die ursprünglich positiv besetzt waren und nach der Gewichtung auf Null reduziert werden. Betrachtet man also die so gewichtete Ist-Tafel von relativen Häufigkeiten, so entspricht diese nicht mehr der Soll-Tafel, sondern die Masse der verschwindenden Sollzellen ist auf die übrigen Zellen gleichmäßig verteilt worden. Dies hat zur Folge,

daß dann auch die gewichteten relativen Häufigkeiten der Ausprägungen einzelner Variablen in der Regel nicht denen der externen Datei entsprechen: Die Abweichungen sind üblicherweise desto gravierender, je mehr Ist-Zellen unbesetzt sind. Die klassische Soll-durch-Ist-Prozedur ist somit unpraktikabel, wenn viele Ist-Zellen des Datensatzes unbesetzt sind.

Man beachte, daß diese Situation sehr leicht eintreten kann: In unserer Untersuchung werden wir vier Variablen mit jeweils 12, 2, 5 bzw. nochmals 5 Ausprägungen zur Adaption benutzen; eine vollständige Tafel besitzt daher ebenfalls bereits $12 \times 2 \times 5 \times 5 = 600$ Zellen. Bei einem Datensatz wie dem ALLBUS 86 mit 3.095 Fällen ist durchaus mit vielen nichtstrukturellen Nullzellen zu rechnen. Dieses Problem verstärkt sich weiter, wenn (wie üblicherweise beim AdM-Redressement) noch mehr Variablen zur Adaption herangezogen werden: Bei der Media-Analyse etwa, für die das AdM-Design ursprünglich konzipiert wurde, werden 6 Variablen mit insgesamt 127596 Ausprägungskombinationen berücksichtigt (Details sind etwa bei Ockelmann (1981) oder Rothe und Wiedenbeck (1987) zu finden). In solchen Fällen kann eine Gewichtung mittels Soll-durch-Ist nicht mehr erfolgen, ohne daß gravierende Abweichungen durch eine überhöhte Zahl von Nullzellen auftreten. In Praxi werden deshalb folgende (ebenfalls rein technisch begründete) Verfahren verwendet:

- Es wird nicht die Gesamttafel adaptiert, sondern nacheinander mehrere Teiltafeln.
- Verschiedene Ausprägungen einer Variablen werden zusammengefaßt.

Die Zusammenfassung (Rekodierung) von Variablen reduziert die Zellenzahl und bei geschickter Anwendung insbesondere auch die Zahl der Nullzellen. In der Praxis (speziell in der Marktforschung etwa bei der Media-Analyse; vgl. hierzu MA-Dokumentationen 1981) wird dieses Verfahren jedoch nur als "Reparaturwerkzeug" verwendet, wenn die standardisierte Redressementprozedur über die Adaption an Teiltafeln wegen dort immer noch auftretenden Nullzellen noch nicht möglich ist. Wir werden dieses zweite Konzept in der vorliegenden Arbeit nicht weiter verfolgen und uns ausschließlich mit der Adaption an Teiltafeln befassen, die im folgenden beschrieben werden sollen.

Aus den zum Redressement zu verwendenden Variablen werden verschiedene Teilmengen ausgewählt, die sich in der Regel überschneiden. Zu jeder Teilmenge

gehört eine eigene Ist- und eine eigene Solltafel (nämlich die jeweiligen Randtafeln der vollständigen Häufigkeitstafeln). Durch die Reduktion auf derartige Randtafeln wird in der Regel das Problem der Nullzellen ganz umgangen oder zumindest wesentlich reduziert.

Die nun durchzuführende Gewichtungsprozedur geht in seinem Grundkonzept auf eine Arbeit von Deming und Stephan (1948) zurück. Sie besteht prinzipiell aus einer iterativen Soll-durch-Ist-Anpassung an alle vorgegebenen Randtafeln. Zunächst wird jedem Fall das Gewicht zugespielt, das sich durch eine Soll-durch-Ist-Adaption an die erste Randtafel ergibt. Die auf diese Weise gewichteten relativen Häufigkeiten der zweiten Tafel werden nun (wiederum mittels Soll-durch-Ist) dem Soll der zweiten Tafel angepaßt. Diese Prozedur wird solange weitergeführt, bis alle Tafeln einmal adaptiert wurden. Danach wird die Gewichtung wiederum mit der ersten Tafel fortgeführt. Es läßt sich zeigen, daß diese iterative Prozedur zu einer Folge von Gewichten führt, die gegen einen Grenzwert konvergiert. Durch Vorgabe eines geeigneten Abbruchkriteriums kann dieser Grenzwert approximiert werden. Das auf diese Weise gewonnene Gewicht besitzt nun folgende Eigenschaften:

- Die Summe der Zellengewichte ist wiederum identisch mit der Gesamtzahl der Fälle
- Jede so gewichtete Randtafel der Adaptionprozedur ist identisch mit der vorgegebenen Soll-Tafel.

Wie bereits angesprochen, ist dieses Konzept aus optischen Gründen ausgesprochen wünschenswert, lassen sich doch hierdurch (werbewirksam) eindimensionale Randverteilungen jeder beliebigen demographischen Variablen im Prinzip durch eine geeignete Gewichtung realisieren. Mehrdimensionale Tafeln werden jedoch nur solange exakt reproduziert, wie sie bei der Adaption auch vollständig einbezogen wurden.

4. Datenbasis und verwendete Gewichtungen. In der vorliegenden Untersuchung soll die Güte unterschiedlicher Gewichtungsvariablen für die Hochrechnung von Daten des ALLBUS 1986 durch einen Vergleich mit den Daten des Mikrozensus 1985 untersucht werden. Die Variablen des ALLBUS-Datensatzes lassen sich hinsichtlich dieser Absicht in drei Kategorien unterteilen. Der aus

wissenschaftlicher Sicht wesentliche Teil ist dabei selbstverständlich diejenige Klasse von Variablen, für die keine Abgleichmöglichkeit mit dem Mikrozensus oder anderen Datenquellen besteht - sie sind ja der Anlaß für die Untersuchung.

Unter denjenigen Variablen, für die eine Abgleichmöglichkeit mit dem Mikrozensus besteht, werden wir noch unterscheiden zwischen Variablen, die bei der Redressementprozedur herangezogen werden und denjenigen, die dazu herangezogen werden, den Effekt der jeweiligen Gewichtung zu überprüfen. Die erste Gruppe besteht aus

- Alter (A)
- Geschlecht (G)
- Bildung (B)
- Stellung im Beruf (S).

Von den übrigen Variablen haben wir die folgenden ausgewählt:

- Familienstand (FAMSTAND)
- Branche (BRANCHE)
- Wochenarbeitsstunden (WARBZEIT)
- Arbeitslosigkeit (ARBLOS)

Diese Größen sind ebenso wie die oben bereits angesprochenen Redressement-Variablen personenbezogen, d.h. sie beziehen sich auf die Person in der Grundgesamtheit und müßten bei Hochrechnungen (im Falle eines exakt realisierten Stichprobenplans) jeweils mit der reduzierten Haushaltsgröße gewichtet werden. Darüber hinaus wurden zusätzlich einige Variablen in den Vergleich mit einbezogen, die sich auf den Haushalt der jeweiligen in die Stichprobe gelangten Personen beziehen, nämlich

- Haushalt mit Kindern (KIND). In dieser Variablen wird erfaßt, ob in dem betreffenden Haushalt Kinder unter 15 Jahren leben oder nicht.
- Haushaltsnettoeinkommen (HHEINK). Neben vier Einkommensklassen werden hierbei die Landwirte als besondere Kategorie erfaßt, da hier die Höhe des Einkommens schwer quantifizierbar ist.

Da es sich (wie bereits angesprochen) bei dem AdM-Design um ein bezüglich der Haushalte selbstgewichtigendes Design handelt, ist bei einer Hochrechnung

von Haushaltsvariablen keine Gewichtung vorzunehmen (immer korrekte Realisation vorausgesetzt). In Tabelle 1 ist für die angesprochenen demographischen Variablen die genaue Abgrenzung der Ausprägungen zu finden. Für diese Ausprägungen enthält Tabelle 1 die im ALLBUS 1986 hierfür tatsächlich realisierten Häufigkeiten sowie die hochgerechneten relativen Häufigkeiten, die sich hieraus ohne Verwendung eines Gewichtes bzw. unter Verwendung des reinen Transformationsgewichtes (reduzierte Haushaltsgröße) ergeben würden sowie schließlich die entsprechenden Häufigkeiten im Mikrozensus des Jahres 1985. Hier sind zwischen den jeweiligen Hochrechnungen und den Zahlen des Mikrozensus zum Teil erhebliche Abweichungen festzustellen. Insbesondere ist zu beobachten, daß oft auch bei personenbezogenen Variablen die Hochrechnung ohne Verwendung einer Gewichtung eher bessere Werte liefert als die theoretisch korrekte Gewichtung mit der reduzierten Haushaltsgröße. Eine mögliche Erklärung hierfür wird sich aus den folgenden Untersuchungen ergeben, wir werden hierauf Abschnitt 5 erneut eingehen.

Für unsere Untersuchung wurden konkret folgende Gewichtungskonzeptionen verwendet:

- (AGSB). Hier wurde ein vollständiges SdI auf die gemeinsame Tafel der zu adaptierenden Variablen A, G, S und B durchgeführt.
- (4 x 3). Mit dieser Kurzform bezeichnen wir im folgenden die Adaption an sämtliche dreidimensionalen Randtafeln, d.h. an die durch Variablengruppen AGS, AGB, ASB und GSB gebildeten Tafeln. Die Adaption erfolgt mit dem oben beschriebenen iterativen Algorithmus und stellt die geringste Vergrößerung im Vergleich zur reinen SdI-Adaption dar.
- (6 x 2). Dieses Kürzel steht für die Adaption an alle sechs zweidimensionalen Randtafeln, d.h. an AG, AS, AB, GS, GB und SB.
- (AG + SB). Hier erfolgt eine Adaption ausschließlich an zwei zweidimensionale Tafeln, nämlich an die Randtafeln AG und BS.
- (AG). Hier erfolgt ausschließlich eine SdI-Anpassung an die Tafel AG.
- (SB). Diese Gewichtung erfolgt analog durch eine reine SdI-Anpassung an die beiden Statusvariablen.

- (A + G + S + B). Diese Gewichtung schließlich resultiert auf einer iterativen Adaption an sämtliche eindimensionalen Randhäufigkeiten.
- (-). Der Vollständigkeit halber bezeichnen wir mit diesem Kürzel den Verzicht auf jegliche externe Anpassung; d.h. die Verwendung einer ungewichteten Hochrechnung.

Die bisher beschriebenen Prozeduren sind grundsätzlich für selbstgewichtende Designs konzipiert, d.h. das Transformationsgewicht wurde bisher in keinem Falle einbezogen. Für die korrekte Gewichtung personenbezogener Daten des ALLBUS ist die reduzierte Haushaltsgröße jedoch zu berücksichtigen. Somit ist die Ist-Tafel bereits eine gewichtete Häufigkeitstafel, wobei die reduzierte Haushaltsgröße als Ausgangsgewicht fungiert. Das nach einer Soll-durch-Ist-Prozedur jeder Person zuzuspielende neue Gewicht ist dann zum Transformationsgewicht in Beziehung zu setzen: Es besteht aus dem Produkt der reduzierten Haushaltsgröße mit dem Soll-durch-Ist-Quotienten der zugehörigen Zelle und einem für alle Fälle gleichen, standardisierenden Normierungsfaktor. Für jede der oben angesprochenen sieben Redressementprozeduren entsteht somit eine zweite durch Einbeziehung der reduzierten Haushaltsgröße in der o.a. Form. Auf diese Weise sind bereits (einschließlich der ungewichteten Hochrechnung und der Transformation) insgesamt 16 verschiedene Hochrechnungsprozeduren entstanden. Unter Verwendung anderer Randtafeln bzw. weiterer Variablen wären prinzipiell beliebig viele weitere Gewichtungsprozeduren möglich. Eine theoretische Analyse der Wirkungsweise dieser Prozeduren ist prinzipiell kaum möglich, mathematisch korrekt sind sie ohnehin nur unter sehr restriktiven Modellannahmen, die in der Regel mit Sicherheit nicht exakt erfüllt sind.

Will man nun die Qualität unterschiedlicher Hochrechnungen anhand der hochgerechneten Häufigkeitsverteilungen einzelner Variablen vergleichen, so ist hierfür ein einsichtiges Qualitätskriterium erforderlich. Tatsächlich ist hier konkret die Ähnlichkeit zweier Häufigkeitsverteilungen zu bestimmen, nämlich der hochgerechneten Häufigkeit einerseits und der tatsächlichen Häufigkeit (als die wir die des Mikrozensus ja immer noch ansehen) andererseits. Als Abweichungsmaß haben wir hier die Chi-Quadrat-Größe, d.h.

$$\chi^2(n,p,q) = n \sum_i (p_i - q_i)^2 / q_i$$

verwendet. Hierbei ist n der Stichprobenumfang (beim ALLBUS '86 $n=3095$) und p und q sind die hochgerechneten bzw. tatsächlichen relativen Häufigkeiten der Ausprägungen.

Diese Zahl ist in unserem Falle ausschließlich als deskriptives Maß interpretierbar, da die Schätzung p ja nicht aus einer Stichprobe vom Umfang n mit unabhängigen, identisch verteilten Realisationen entstanden ist. Eine große Abweichung im Sinne des Chi-Quadrat-Tests kann hier also nicht sofort als Argument gegen eine bestimmten Gewichtungsprozedur verwendet werden. Die Maßzahl erlaubt lediglich einen Vergleich von hochgerechneten Häufigkeitsverteilung einer Variablen zwischen den verschiedenen Gewichtungsprozeduren. Ihre absolute Größe ist nicht aussagekräftig.

Es sollte an dieser Stelle allerdings nicht verschwiegen werden, daß prinzipiell auch Stichprobenfehler im Mikrozensus auftreten können. Auch hier wird nämlich eine Art Redressementprozedur durchgeführt, um die Daten des Mikrozensus zeitreihentechnisch zu "glätten", d.h. Daten werden denen des Vorjahres in etwa angepaßt, um zufallsbedingte Sprünge zu eliminieren. Es werden hierzu jedoch ausschließlich ganzzahlige Gewichte verwendet, d.h. die Prozedur entspricht dem klassischen "Karten entfernen/Karten doppeln"-Verfahren der Vorcomputerzeit. Die auf diese Weise erfolgte Fortschreibung der Volkszählung 1970 bis zum Jahre 1985 hat mit großer Wahrscheinlichkeit ebenfalls Verzerrungen zur Folge. Die bisher verfügbaren Daten der Volkszählung 1987 scheinen dies zu bestätigen. Derart extreme Abweichungen zwischen Mikrozensus und ALLBUS jedoch, wie sie in den Variablen "Bildung" und "Stellung im Beruf" tatsächlich feststellbar sind, lassen sich auf diese Weise mit Sicherheit nicht erklären.

5. Durchführung und Ergebnisse. Wie im vorigen Abschnitt beschrieben, lassen sich die in dieser Arbeit verwendeten Gewichte jeweils unter Verwendung des Deming-Stephan-Algorithmus bestimmen, der in der Literatur auch IPF-Algorithmus (für "iterative proportional fitting") genannt wird. Vorwiegend zur Anwendung gelangt dieser IPF-Algorithmus bei der Schätzung von Parametern im hierarchischen log-linearen Modell. In diesen Fällen ist die Starttafel eine mit dem Modell kompatible Tafel, im einfachsten Fall eine Tafel, bei der alle Zellen identisch besetzt sind. Die verfügbaren Programmpakete wie ECTA (ebenso wie die Implementation von ECTA im Programm CDAS) oder die Prozedur

HILOGLINEAR im SPSS verfügen nicht über die Option, beliebige Starttafeln einzugeben; dies wäre aber zur Gewichtskonstruktion in unserem Sinne erforderlich. Aus diesem Grund wurde eine Version des IPF-Algorithmus für beliebige Startwerte im GAUSS, Ver. 2.0, realisiert. Es wurde eine Library erstellt, die alle erforderlichen Prozeduren umfaßt; diese Library mit der Dokumentation der enthaltenen Prozeduren findet sich im Anhang zu dieser Arbeit. GAUSS hat leider den Nachteil, daß als Datenstrukturen nur Zeichenketten und zweidimensionale Felder zulässig sind, mehrdimensionale Tafeln somit nicht einfach darstellbar sind. In der vorliegenden Version wurden die Zellbesetzungszahlen somit jeweils lexikographisch in einem Spaltenvektor angeordnet; für die Bestimmung von Randsummen, wie sie für die Adaption an Randtafeln erforderlich sind, wurden dann geeignete Matrixmanipulationen als Prozeduren herangezogen. Dieser Nachteil von GAUSS wird durch die bekannt hohe Rechengeschwindigkeit und die große Rechengenauigkeit dieses Pakets wieder wettgemacht.

Für alle im vorigen Abschnitt angesprochenen Variablen wurde die mit der jeweiligen Gewichtung bestimmte Hochrechnung der Häufigkeitsverteilung ermittelt. Diese Häufigkeitsverteilungen finden sich in Tabelle 2.

Anhand des beschriebenen Kriteriums (Chi-Quadrat-Abstand) kann nun die Qualität der Hochrechnung und damit der Gewichtungskonzeption diskutiert werden. Hierzu wurden für jede der in Tab.2 konstruierten Häufigkeitsverteilungen der Chi-Quadrat-Abstand zur Häufigkeit im Mikrozensus bestimmt. Diese Zahlen finden sich in Tabelle 3.

Wir betrachten zunächst die 4 Variablen, die bei der Gewichtungsprozedur berücksichtigt wurden (A, G, S und B). Hier sei zunächst darauf hingewiesen, daß durch die Gewichtungsprozedur nur die Zellhäufigkeiten der jeweils berücksichtigten Randtafeln an die entsprechenden Tafeln des Solls adaptiert wurden. Die Einbeziehung des Transformationsgewichts (REDHH) wirkt sich erst bei der Vergabe der Gewichte bei den Einzelfällen aus. Die gewichteten Häufigkeiten der bei der Gewichtung berücksichtigten Randtafeln sind dann jeweils in beiden Fällen identisch. Dies hat zur Folge, daß die gewichteten Hochrechnungen eindimensionaler Ränder mit und ohne Transformationsgewicht die gleiche Häufigkeitsverteilung liefern - vorge setzt, daß es sich dabei um eine bei der Gewichtung berücksichtigte Variable handelt. So ist beispielsweise die Häufigkeitsverteilung der Variablen S etwa bei der Verwendung des

Gewichtes (AG + SB) und damit auch der Chi-Quadrat-Abstand zur Verteilung des Mikrozensus bei der Hochrechnung mit und ohne Transformation identisch, nicht jedoch bei Verwendung des Gewichtes (AG), da hier die Variable S bei der Adaption nicht einbezogen wurde. Die Abweichung von Null beim Chi-Quadrat-Abstand zum Mikrozensus sind immer dann sehr gering, wenn die jeweilige Variable bei der Gewichtung einbezogen wurde. Daß der Abstand überhaupt von Null abweicht, ist durch den in Abschnitt 3 beschriebenen "Massenverlust" zu erklären, der immer dann entsteht, wenn einige Ist-Zellen unbesetzt sind, die zugehörigen Soll-Zellen jedoch nicht. Interessant sind auch hier vorwiegend diejenigen Gewichtungen, bei denen die jeweilige Variable nicht einbezogen wurde (bei S also speziell die Gewichte (-) und (AG), jeweils mit und ohne Berücksichtigung der reduzierten Haushaltsgröße).

Wie bereits früher erwähnt, sind bei einigen Variablen (konkret: bei A, G und B) die ungewichteten Hochrechnungen besser als Hochrechnungen unter Verwendung ausschließlich des reinen Transformationsgewichts. In diesen Fällen ist allerdings eine Verbesserung der Hochrechnung durch eine Gewichtung festzustellen, selbst wenn die jeweilige Variable nicht einbezogen wurde. Bei der Variablen S dagegen ist die Situation umgekehrt: Hier ist die Verwendung des Transformationsgewichts besser als das naive Vorgehen, dafür führt jedoch eine Adaption an Alter und Geschlecht sogar zu einer deutlichen Verschlechterung der Hochrechnung (Chi-Quadrat-Wert von 63,7 gegenüber 31,4 ohne Redressement).

Ähnlich heterogen ist die Situation beim Vergleich der übrigen personenbezogenen demographischen Variablen: Die (theoretisch falsche) ungewichtete Hochrechnung führt bei FAMSTAND und bei ARBLOS zu besseren Schätzungen als die reine Verwendung des Transformationsgewichts. Bei den Variablen BRANCHE und WARBEIT ist die einfache Hochrechnung mit dem Transformationsgewicht zwar besser, jedoch wird durch die Einbeziehung anderer Adaptionen an A, G, S und B höchstens noch eine geringfügige weitere Verbesserung erreicht. Besonders bemerkenswert an den Gewichtungen zur Variablen ARBLOS und FAMSTAND ist die Tatsache, daß sämtliche verwendeten Gewichtungen deutlich schlechtere Anpassungen liefern als die "naive" Hochrechnung ohne jegliches Gewicht.

Bei den beiden in die Analyse einbezogenen Haushaltsvariablen ist das Bild wiederum völlig unterschiedlich: Für HHEINK liefert die (theoretisch korrekte) einfache Hochrechnung die beste Schätzung, fast alle externen Adaptionen an A, G, S und B sind schlechter. Die Variable KIND ist dagegen die einzige Variable, die sich wie "erwünscht" verhält: Die ungewichtete Hochrechnung ist besser als die mit dem Transformationsgewicht versehene (wie es sein sollte); die Verwendung der verschiedenen Adaptionsgewichte führt darüberhinaus zu einer deutlichen Verbesserung der Hochrechnung. Dies ist allerdings nicht sonderlich überraschend: Kinder unter 15 Jahren sind vorwiegend in den Altersgruppen zwischen 20 und 40 Jahren anzutreffen, damit ist die Variable sehr eng an die Altersvariable gekoppelt: Die Adaption ist immer dann besonders gut, wenn diese Variable bei der Gewichtung einbezogen wurde.

Die oben beschriebenen Effekte der Gewichtungsprozedur auf die Hochrechnung von Verteilungen liefern ein außerordentlich heterogenes und damit entmutigendes Bild. Obwohl entsprechende Untersuchungen bisher nicht angestellt wurden, ist zu erwarten, daß die unübersichtlichen Auswirkungen von Gewichtungen sich eher verstärkt zeigen, wenn zusätzlich das Verhalten der Hochrechnung für die gemeinsame Verteilung zweier oder mehrerer Variablen untersucht wird.

Erklärungen für das beobachtete Verhalten lassen sich aus den vorliegenden Zahlen nicht direkt ableiten; für zwei der auffälligsten Effekte sei jedoch an dieser Stelle eine Vermutung geäußert:

- (1) Bei den meisten personenbezogenen Variablen erwies sich die ungewichtete Hochrechnung (obwohl theoretisch falsch) als besser als die Hochrechnung unter Verwendung des theoretisch begründeten Transformationsgewichts. Offenbar hebt sich der "Fehler", der bei der Verwendung der naiven Hochrechnung gemacht wird, gegen den Fehler, der aus mangelhafter Realisierung des Designs resultiert, auf: Es ist zu vermuten, daß kleine Haushalte, spezielle Haushalte mit einer reduzierten Haushaltsgröße von 1, schlechter erreichbar und damit in der Stichprobe unterrepräsentiert sind (also eine kleinere Chance haben, in die Stichprobe zu kommen, als große Haushalte). Bei der naiven Hochrechnung erhalten sie allerdings ein "zu hohes" Gewicht, das möglicherweise in manchen Fällen die falsche Repräsentativität wieder ausgleicht. Dieser Effekt würde sich insbesondere bei Variablen bemerkbar machen, die in irgendeiner Weise mit

der Haushaltsgröße korrelieren. Dies ist z.B. bekannt für die Variablen Alter und Geschlecht (jüngere Singles, alleinstehende ältere Frauen etc.).

- (2) Für die schlechten Ergebnisse bei der Hochrechnung des Anteils der Arbeitslosen nach Adaption an Statusvariablen gibt es eine sehr einleuchtende Erklärung. Aufgrund von Untersuchungen am sozio-ökonomischen Panel (Rendtel 1989) kann davon ausgegangen werden, daß bei Arbeitslosen die Bereitschaft zur Teilnahme an einer Befragung deutlich reduziert ist. Ferner ist bekannt, daß Arbeitslose insbesondere in der Gruppe der Personen ohne Schulabschluß (d.h. in der Gruppe $B = 1$) besonders stark vertreten sind. Damit liegt der Anteil der Arbeitslosen in der Gruppe $B = 1$ der Stichprobe deutlich unter dem tatsächlichen Wert für diese Gruppe. Gleichzeitig ist aber auch die Gruppe $B = 1$ in der Gesamtstichprobe unterrepräsentiert; durch eine Adaption an Statusvariablen erhält diese Gruppe in der Stichprobe zusätzlich ein erhöhtes Gewicht. Je höher der Anteil der Verweigerer in der Gruppe $B = 1$ ist, desto mehr wird die durch die Verweigerung bedingte Verzerrung in der Hochrechnung durch die Adaption zusätzlich verstärkt. Dies führt schließlich zu der beobachteten Verschlechterung des Chi-Quadrat-Wertes.

6. Schlußbemerkungen. In der vorliegenden Arbeit wurden nur Variablen untersucht, deren tatsächliche Verteilung sich zumindest in etwa anhand des Mikrozensus feststellen ließ. Bei anderen Variablen lassen sich zwar auch Unterschiede in der Hochrechnung durch Verwendung der Gewichte feststellen; aufgrund eines Vergleichsmaßstabs fehlt jedoch in diesen Fällen die Möglichkeit, Aussagen über die Anpassungsgüte der Gewichtung zu treffen. Es muß jedoch erwartet werden, daß die unerfreulichen Effekte, die sich bei der Analyse der demographischen Variablen gezeigt haben, auch für solche Variablen Gültigkeit haben. Es erscheint daher schwerer denn je, spezielle Hochrechnungsprozeduren zu empfehlen. Die Verwendung von Globalgewichten, die zumindest auffälligste Diskrepanzen zum Mikrozensus "ausbügeln" sollen, erscheint nicht mehr angebracht. Es bleibt prinzipiell keine andere Möglichkeit, als in jedem einzelnen Fall einer Hochrechnung zunächst zu diskutieren, inwieweit der Ausfallmechanismus (Erreichbarkeit, Verweigerung etc.) mit der hochzurechnenden Variablen in Verbindung stehen könnte. Eine Verbesserung der Hochrechnung durch Adaption externer Variablen ist nur in solchen Fällen möglich, in denen sich rechtfertigen läßt, daß diese Variablen die

Verbindung zwischen Ausfallgeschehen und der eigentlich zu analysierenden Variablen kontrollieren. Dies bedeutet auch, daß für jede einzelne Analyse die Frage nach der geeigneten Gewichtungsvariablen neu gestellt werden muß, d.h. daß es keine globalen Gewichtungsvariablen geben kann. Will man auf solche Überlegungen verzichten, ist die Verwendung keines Gewichts ebenso gut oder schlecht wie die jedes anderen.

Literatur:

Deming, W.E. / Stefan, F.F., 1940: On a least square adjustment of a sampled frequency table when the expected marginal totals are known.

Ann. Math. Statist. **11**, 427-444.

Kirschner, H.-P. (1984). ALLBUS 1980: Stichprobenplan und Gewichtung.

In: K.-U. Mayer/P.Schmidt (Hrsg.). Allgemeine Bevölkerungsumfrage der Sozialwissenschaften. ZUMA-Monographien Band 5, Frankfurt.

Ockelmann, E. (1981). Das Redressement der Media-Analyse als analytische Korrektur der Feldergebnisse.

In: AG.MA (Hrsg.), MA 81-Dokumentationen, S. 93-106.

Rendtel, U. (1989). Panelmortalität: Eine Analyse der Antwortausfälle beim Sozio-ökonomischen Panel.

Vortrag, 2. Sitzung der Arbeitsgruppe "Auswertung von Panel-Daten", Mannheim, Januar 1989.

Rothe, G. / Wiedenbeck, M. (1987). Stichprobengewichtung: Ist Repräsentativität machbar?

ZUMA-Nachrichten **21**, 43-58.

Tab.1: Relative Häufigkeiten (in %) der Ausprägungen abgleichbarer Variablen aus ALLBUS '86 und Mikrozensus '85.

Variable bzw. Ausprägung	Schätzung aus ALLBUS '86		Mikrozensus 1985
	ohne Transformationsgewicht	mit	

a) Alter

(A)

Bis 19	3.13	4.43	4.30
20 bis 24	9.76	10.92	10.30
25 bis 29	10.79	9.92	8.90
30 bis 34	9.40	8.73	8.00
35 bis 39	9.69	9.43	7.10
40 bis 44	9.89	10.30	8.20
45 bis 49	9.21	10.61	9.80
50 bis 54	6.72	8.08	7.90
55 bis 59	6.82	7.22	8.00
60 bis 64	8.24	7.85	7.60
65 bis 69	5.59	4.58	4.70
70 und älter	10.76	7.93	15.20

b) Geschlecht

(G)

männlich	46.69	49.21	46.40
weiblich	53.31	50.79	53.60

c) Bildung

(B)

Hauptschule ohne Lehre	19.58	18.69	31.70
Hauptschule mit Lehre	38.45	39.29	37.30
Realschule	23.17	23.28	18.30
Abitur	10.27	10.63	6.50
Hochschulabschluß	8.53	8.11	6.20

d) Stellung im Beruf

(S)

Selbständige	5.49	6.31	6.20
Beamte	6.24	6.34	5.10
Angestellte	22.00	22.16	21.20
Arbeiter	14.44	15.32	18.60
Nicht erwerbstätig	51.83	49.87	48.80

noch Tab.1: Relative Häufigkeiten der Ausprägungen abgleichbarer Variablen
aus ALLBUS '86 und Mikrozensus '85

Variable bzw. Ausprägung	Schätzung aus ALLBUS '86		Mikrozensus 1985
	ohne Transformationsgewicht	mit	

e) Familienstand
(FAMSTAND)

Ledig	21.49	21.11	22.90
Verheiratet	61.52	68.40	61.00
Geschieden/Verwitwet	17.00	10.49	16.00

e) Branche
(BRANCHE)

Land/Forst/Tier/Fisch	1.20	1.68	2.50
Energie/Wasser/Bergbau	1.03	1.08	1.00
Chemie/Kunststoff/Stein	2.36	2.53	2.50
Eisen/Metall/Stahl	1.91	2.10	1.60
Stahl/Masch./KFZ-Bau	4.36	4.46	5.00
El.Tech./Feinmech./Optik	3.30	3.23	3.10
Holz/Papier/Druck	1.23	1.40	1.70
Leder/Textil/Bekleidung	1.03	1.04	1.30
Nahrung/Genußmittel	1.55	1.84	1.40
Bau/Ausbau/Bauhilfe	2.75	3.09	3.00
Handel: Groß/Einzel/Verk.	5.20	5.42	6.40
Bahn/Post/Verkehr	3.00	3.19	3.00
Kredit/Finanz/Vers.	1.97	2.04	1.90
Dienstl./Priv. HH	4.85	4.91	10.40
Pol./priv. Organis.	11.47	11.27	6.30
Nicht erwerbstätig	52.79	50.73	48.80

f) Wochenarbeitsstunden
(WARBZEIT)

Bis 20	3.30	3.47	3.30
21 bis 30	3.07	3.05	2.60
31 bis 35	0.71	0.72	0.60
36 bis 39	4.78	4.82	5.20
40	19.19	19.98	30.70
41 bis 42	2.39	2.46	0.80
43 bis 45	4.46	4.44	1.30
46 bis 50	4.75	4.88	2.30
51 bis 60	4.01	4.30	2.40
61 und mehr	1.78	2.20	2.00
Nicht erwerbstätig	51.57	49.68	48.80

noch Tab.1: Relative Häufigkeiten der Ausprägungen abgleichbarer Variablen
aus ALLBUS '86 und Mikrozensus '85

Variable bzw. Ausprägung	Schätzung aus ALLBUS '86		Mikrozensus 1985
	ohne Transformationsgewicht	mit	

g) Arbeitslosigkeit
(ARBLOS)

Ja	3.59	3.20	3.70
Nein	96.41	96.80	96.30

h) Haushalt mit Kindern unter 15 Jahren
(KIND)

Ja	27.53	28.63	23.60
Nein	72.47	71.37	76.40

i) Haushaltsnettoeinkommen
(HHEINK)

bis 999 DM	6.33	3.89	5.80
1.000 - 2.999 DM	49.05	44.42	47.90
3.000 - 4.999 DM	25.46	29.70	25.20
5.000 und mehr DM	6.30	7.93	7.70
Landwirte	12.86	14.06	13.50

Tab.2: Hochgerechnete relative Häufigkeiten der Variablen aus Tab.1
a) ohne Transformation.

Variable:	A									
		-	AGBS	4x3	6x2	AG+SB	BS	AG	A+G +B+S	MZ85
Bis 19		3.13	3.37	3.50	4.29	4.29	3.79	4.29	4.29	4.30
20 bis 24		9.76	10.46	10.33	10.32	10.32	8.53	10.32	10.32	10.30
25 bis 29		10.79	8.98	9.04	8.87	8.87	10.05	8.87	8.87	8.90
30 bis 34		9.40	8.04	8.05	7.99	7.99	8.71	7.99	7.99	8.00
35 bis 39		9.69	7.16	7.22	7.11	7.11	9.56	7.11	7.11	7.10
40 bis 44		9.89	8.24	8.37	8.22	8.22	9.65	8.22	8.22	8.20
45 bis 49		9.21	10.04	10.23	9.81	9.81	10.11	9.81	9.81	9.80
50 bis 54		6.72	7.87	8.00	7.85	7.85	7.30	7.85	7.85	7.90
55 bis 59		6.82	8.04	8.08	8.03	8.03	7.44	8.03	8.03	8.00
60 bis 64		8.24	7.58	7.59	7.57	7.57	8.23	7.57	7.57	7.60
65 bis 69		5.59	4.71	4.56	4.73	4.73	5.56	4.73	4.73	4.70
70 und älter		10.76	15.51	15.04	15.23	15.23	11.06	15.23	15.23	15.20

Variable:	G								
	-	AGBS	4 x 3	6 x 2	AG+SB	BS	AG	A+G +B+S	MZ85
männlich	46.69	45.92	46.40	46.40	46.40	45.65	46.40	46.40	46.40
weiblich	53.31	54.08	53.60	53.60	53.60	54.35	53.60	53.60	53.60

Variable:	B								
	-	AGBS	4 x 3	6 x 2	AG+SB	BS	AG	A+G +B+S	MZ85
Hauptsch. o. Lehre	19.58	31.64	31.66	31.69	31.67	31.67	22.01	31.67	31.70
Hauptsch. m. Lehre	38.45	37.52	37.30	37.30	37.30	37.30	37.01	37.30	37.30
Realschule	23.17	18.52	18.32	18.30	18.32	18.32	22.89	18.32	18.30
Abitur	10.27	6.43	6.54	6.54	6.54	6.54	10.21	6.54	6.50
Hochschulabschluß	8.53	5.90	6.17	6.17	6.17	6.17	7.88	6.17	6.20

Variable:	S									
		-	AGBS	4 x 3	6 x 2	AG+SB	BS	AG	A+G +B+S	MZ85
Selbständige	5.49	5.19	6.23	6.21	6.23	6.23	5.16	6.23	6.20	
Beamte	6.24	4.80	5.11	5.12	5.12	5.12	5.89	5.12	5.10	
Angestellte	22.00	21.37	21.25	21.18	21.25	21.25	20.66	21.25	21.20	
Arbeiter	14.44	18.56	18.61	18.62	18.61	18.61	14.30	18.61	18.60	
Nicht erwerbstätig	51.83	50.08	48.80	48.87	48.79	48.79	53.99	48.79	48.80	

Tab.2: Hochgerechnete relative Häufigkeiten der Variablen aus Tab. 1
b) mit Transformation.

Variable:	A								
	-	AGBS	4 x 3	6 x 2	AG+SB	BS	AG	A+G +B+S	MZ85
Bis 19	4.43	3.37	3.50	4.29	4.29	5.42	4.29	4.29	4.30
20 bis 24	10.92	10.46	10.33	10.32	10.32	9.18	10.32	10.32	10.30
25 bis 29	9.92	8.98	9.04	8.87	8.87	9.16	8.87	8.87	8.90
30 bis 34	8.73	8.04	8.05	7.99	7.99	8.05	7.99	7.99	8.00
35 bis 39	9.43	7.16	7.22	7.11	7.11	9.08	7.11	7.11	7.10
40 bis 44	10.30	8.24	8.37	8.22	8.22	9.85	8.22	8.22	8.20
45 bis 49	10.61	10.04	10.23	9.81	9.81	11.43	9.81	9.81	9.80
50 bis 54	8.08	7.87	8.00	7.85	7.85	8.79	7.85	7.85	7.90
55 bis 59	7.22	8.04	8.08	8.03	8.03	7.79	8.03	8.03	8.00
60 bis 64	7.85	7.58	7.59	7.57	7.57	8.11	7.57	7.57	7.60
65 bis 69	4.58	4.71	4.56	4.73	4.73	4.71	4.73	4.73	4.70
70 und älter	7.93	15.51	15.04	15.23	15.23	8.46	15.23	15.23	15.20

Variable:	G								
	-	AGBS	4 x 3	6 x 2	AG+SB	BS	AG	A+G +B+S	MZ85
männlich	49.21	45.92	46.40	46.40	46.40	47.31	46.40	46.40	46.40
weiblich	50.79	54.08	53.60	53.60	53.60	52.69	53.60	53.60	53.60

Variable:	B								
	-	AGBS	4 x 3	6 x 2	AG+SB	BS	AG	A+G +B+S	MZ85
Hauptsch. o. Lehre	18.69	31.64	31.66	31.69	31.67	31.67	22.15	31.67	31.70
Hauptsch. m. Lehre	39.29	37.52	37.30	37.30	37.30	37.30	37.41	37.30	37.30
Realschule	23.28	18.52	18.32	18.30	18.32	18.32	22.68	18.32	18.30
Abitur	10.63	6.43	6.54	6.54	6.54	6.54	10.23	6.54	6.50
Hochschulabschluß	8.11	5.90	6.17	6.17	6.17	6.17	7.53	6.17	6.20

Variable:	S								
	-	AGBS	4 x 3	6 x 2	AG+SB	BS	AG	A+G +B+S	MZ85
Selbständige	6.31	5.19	6.23	6.21	6.23	6.23	5.81	6.23	6.20
Beamte	6.34	4.80	5.11	5.12	5.12	5.12	5.77	5.12	5.10
Angestellte	22.16	21.37	21.25	21.18	21.25	21.25	20.11	21.25	21.20
Arbeiter	15.32	18.56	18.61	18.62	18.61	18.61	14.19	18.61	18.60
Nicht erwerbstätig	49.87	50.08	48.80	48.87	48.79	48.79	54.11	48.79	48.80

Tab.2: Hochgerechnete relative Häufigkeiten der Variablen aus Tab. 1 (Fort.)
a) ohne Transformation.

Variable: FAMSTAN

	-	AGBS	4 x 3	6 x 2	AG+SB	BS	AG	A+G +B+S	MZ85
Ledig	21.49	20.94	21.10	22.10	21.69	20.18	22.40	21.51	22.90
Verheiratet	61.52	59.31	59.48	58.21	58.66	61.57	58.18	58.72	61.00
Gesch./Verw.	17.00	19.75	19.42	19.70	19.65	18.25	19.42	19.77	16.00

Variable: BRANCHE

	-	AGBS	4 x 3	6 x 2	AG+SB	BS	AG	A+G +B+S	MZ85
Land/...	1.20	1.56	1.77	1.84	1.78	1.74	1.13	1.52	2.50
Energie ...	1.03	1.24	1.29	1.16	1.17	1.17	1.03	1.13	1.00
Chemie/...	2.36	2.88	2.78	2.75	2.70	2.74	2.29	2.70	2.50
Eisen/...	1.91	2.00	2.04	1.99	2.14	2.22	1.82	2.09	1.60
Stahl/...	4.36	4.80	4.59	4.63	4.70	4.71	4.21	4.67	5.00
El.Tech./...	3.30	3.19	3.23	3.27	3.40	3.44	3.16	3.44	3.10
Holz/...	1.23	1.55	1.62	1.60	1.56	1.57	1.20	1.53	1.70
Leder/...	1.03	1.21	1.31	1.39	1.28	1.23	1.03	1.35	1.30
Nahrung/...	1.55	1.89	2.00	1.84	1.77	1.79	1.51	1.84	1.40
Bau/...	2.75	2.94	3.00	3.10	3.11	3.14	2.72	3.12	3.00
Handel	5.20	5.33	5.59	5.75	5.60	5.44	5.00	5.64	6.40
Bahn/...	3.00	3.00	3.27	3.51	3.34	3.32	2.94	3.20	3.00
Kredit/...	1.97	2.06	2.03	1.92	1.90	1.93	1.85	1.88	1.90
Dienstl./...	4.85	4.92	5.30	5.17	5.18	5.14	4.63	5.25	10.40
Pol./...	11.47	10.15	10.24	10.01	10.24	10.44	10.61	10.49	6.30
Nicht erwerbstätig	52.79	51.29	49.94	50.06	50.13	49.99	54.88	50.15	48.80

Variable: WARBZEI

	-	AGBS	4 x 3	6 x 2	AG+SB	BS	AG	A+G +B+S	MZ85
Bis 20	3.30	3.48	3.45	3.41	3.42	3.52	2.88	3.49	3.30
21 bis 30	3.07	3.34	3.87	3.55	3.35	3.33	2.82	3.29	2.60
31 bis 35	0.71	0.59	0.63	0.65	0.68	0.71	0.63	0.70	0.60
36 bis 39	4.78	5.12	5.14	5.22	5.22	5.24	4.57	5.22	5.20
40	19.19	20.80	21.05	21.09	21.04	20.55	18.85	20.87	30.70
41 bis 42	2.39	2.38	2.25	2.31	2.34	2.41	2.34	2.39	0.80
43 bis 45	4.46	4.67	4.68	4.60	4.55	4.51	4.28	4.47	1.30
46 bis 50	4.75	4.56	4.64	4.53	4.77	5.00	4.39	4.78	2.30
51 bis 60	4.01	3.51	3.82	3.86	3.86	3.99	3.84	3.94	2.40
61 und mehr	1.78	1.71	1.91	2.00	1.94	1.99	1.69	1.96	2.00
Nicht erwerbstätig	51.57	49.84	48.57	48.80	48.83	48.75	53.71	48.89	48.80

Tab.2: Hochgerechnete relative Häufigkeiten der Variablen aus Tab. 1 (Fort.)
b) mit Transformation.

Variable: FAMSTAN

	-	AGBS	4 x 3	6 x 2	AG+SB	BS	AG	A+G +B+S	MZ85
Ledig	21.11	19.62	19.76	20.46	19.95	19.77	20.44	19.63	22.90
Verheiratet	68.40	65.30	65.45	64.63	65.30	68.55	65.01	65.98	61.00
Gesch./Verw.	10.49	15.08	14.79	14.90	14.75	11.68	14.56	14.39	16.00

Variable: BRANCHE

	-	AGBS	4 x 3	6 x 2	AG+SB	BS	AG	A+G +B+S	MZ85
Land/...	1.68	1.63	1.85	1.99	2.03	1.93	1.55	1.82	2.50
Energie ...	1.08	1.19	1.24	1.14	1.19	1.18	1.04	1.11	1.00
Chemie/...	2.53	2.98	2.92	2.76	2.75	2.78	2.33	2.66	2.50
Eisen/...	2.10	2.06	2.09	2.05	2.24	2.27	1.96	2.19	1.60
Stahl/...	4.46	4.76	4.58	4.55	4.65	4.68	4.08	4.58	5.00
El.Tech./...	3.23	3.13	3.13	3.19	3.24	3.28	2.96	3.30	3.10
Holz/...	1.40	1.64	1.70	1.75	1.62	1.68	1.29	1.63	1.70
Leder/...	1.04	1.14	1.22	1.27	1.19	1.16	0.96	1.29	1.30
Nahrung/...	1.84	2.08	2.23	2.10	2.03	2.03	1.67	2.15	1.40
Bau/...	3.09	2.93	3.01	3.14	3.19	3.29	2.85	3.16	3.00
Handel	5.42	5.31	5.57	5.66	5.49	5.45	4.89	5.58	6.40
Bahn/...	3.19	3.11	3.41	3.59	3.42	3.44	2.93	3.22	3.00
Kredit/...	2.04	2.01	1.97	1.92	1.92	1.90	1.88	1.85	1.90
Dienstl./...	4.91	4.79	5.20	5.09	5.07	4.98	4.62	5.10	10.40
Pol./...	11.27	10.01	10.03	9.85	9.97	10.13	10.10	10.33	6.30
Nicht erwerbstätig	50.73	51.20	49.87	49.97	50.00	49.83	54.88	50.05	48.80

Variable: WARBEI

	-	AGBS	4 x 3	6 x 2	AG+SB	BS	AG	A+G +B+S	MZ85
Bis 20	3.47	3.67	3.66	3.64	3.60	3.64	2.94	3.87	3.30
21 bis 30	3.05	3.39	3.89	3.57	3.23	3.17	2.71	3.37	2.60
31 bis 35	0.72	0.62	0.66	0.68	0.71	0.72	0.62	0.76	0.60
36 bis 39	4.82	5.07	5.10	5.01	5.13	5.13	4.38	5.16	5.20
40	19.98	20.52	20.75	21.08	20.93	20.71	18.49	20.58	30.70
41 bis 42	2.46	2.38	2.23	2.25	2.30	2.41	2.28	2.31	0.80
43 bis 45	4.44	4.52	4.55	4.39	4.44	4.40	4.14	4.27	1.30
46 bis 50	4.88	4.60	4.68	4.54	4.86	4.93	4.51	4.72	2.30
51 bis 60	4.30	3.55	3.88	3.86	3.84	3.98	3.98	3.87	2.40
61 und mehr	2.20	1.68	1.91	2.03	2.00	2.13	1.99	2.07	2.00
Nicht erwerbstätig	49.68	49.97	48.69	48.95	48.95	48.78	53.95	49.01	48.80

Tab.2: Hochgerechnete relative Häufigkeiten der Variablen aus Tab. 1 (Fort.)
a) ohne Transformation.

Variable: ARBLOS

	-	AGBS	4 x 3	6 x 2	AG+SB	BS	AG	A+G +B+S	MZ85
ja	3.59	2.78	2.95	2.75	2.67	3.06	3.35	2.72	3.70
nein	96.41	97.22	97.05	97.25	97.33	96.94	96.65	97.28	96.30

Variable: KIND

	-	AGBS	4 x 3	6 x 2	AG+SB	BS	AG	A+G +B+S	MZ85
ja	27.53	23.01	23.22	23.14	23.50	26.94	23.50	23.64	23.60
nein	72.47	76.99	76.78	76.86	76.50	73.06	76.50	76.36	76.40

Variable: HHEINK

	-	AGBS	4 x 3	6 x 2	AG+SB	BS	AG	A+G +B+S	MZ85
bis 999 DM	6.33	6.97	6.89	7.40	6.56	6.30	6.92	6.60	5.80
1.000 - 2.999 DM	49.05	51.94	51.69	51.02	50.76	50.50	49.59	50.85	47.90
3.000 - 4.999 DM	25.46	23.35	22.90	23.16	23.96	24.61	24.51	24.13	25.20
5.000 und mehr DM	6.30	5.25	5.55	5.58	5.43	5.47	6.01	5.47	7.70
Landwirte	12.86	12.48	12.98	12.84	13.30	13.12	12.97	12.95	13.50

Tab.2: Hochgerechnete relative Häufigkeiten der Variablen aus Tab. 1 (Fort.)
b) mit Transformation.

Variable: ARBLOS

	-	AGBS	4 x 3	6 x 2	AG+SB	BS	AG	A+G +B+S	MZ85
ja	3.20	2.49	2.67	2.49	2.31	2.88	2.90	2.28	3.70
nein	96.80	97.51	97.33	97.51	97.69	97.12	97.10	97.72	96.30

Variable: KIND

	-	AGBS	4 x 3	6 x 2	AG+SB	BS	AG	A+G +B+S	MZ85
ja	28.63	23.74	23.99	23.96	24.35	27.70	24.54	24.47	23.60
nein	71.37	76.26	76.01	76.04	75.65	72.30	75.46	75.53	76.40

Variable: HHEINK

	-	AGBS	4 x 3	6 x 2	AG+SB	BS	AG	A+G +B+S	MZ85
bis 999 DM	3.89	4.96	4.91	5.19	4.48	4.02	4.66	4.39	5.80
1.000 - 2.999 DM	44.42	48.91	48.71	47.74	47.25	46.36	46.14	47.16	47.90
3.000 - 4.999 DM	29.70	26.79	26.14	26.82	27.72	28.80	28.04	27.99	25.20
5.000 und mehr DM	7.93	6.22	6.51	6.66	6.52	6.78	7.20	6.68	7.70
Landwirte	14.06	13.12	13.73	13.58	14.04	14.04	13.96	13.77	13.50

Tab.3: Chi-Quadrat-Abstände der Verteilungen aus Tab.2 zum Mikrozensus.

		A	G	B	S		
o h n e	(-)	138.0439	0.1033	327.5132	52.7366		
	(AGBS)	8.4752	0.2911	0.6154	7.6986		
	(4 x 3)	6.7578	0.0000	0.0134	0.0061		
	(6 x 2)	0.0301	0.0000	0.0135	0.0035		
	(AG+SB)	0.0301	0.0000	0.0124	0.0079		
	(AG)	0.0301	0.0000	213.2343	65.7096		
	(BS)	102.4287	0.7110	0.0124	0.0079		
Transfor- mation	(A+G+B+S)	0.0301	0.0000	0.0124	0.0079		
m i t	(-)	248.9662	9.7594	379.6788	31.3824		
	(AGBS)	8.4752	0.2911	0.6154	7.6986		
	(4 x 3)	6.7569	0.0000	0.0134	0.0061		
	(6 x 2)	0.0301	0.0000	0.0135	0.0035		
	(AG+SB)	0.0301	0.0000	0.0124	0.0079		
	(AG)	0.0301	0.0000	203.1123	63.7271		
	(BS)	210.9339	1.0271	0.0124	0.0079		
Transfor- mation	(A+G+B+S)	0.0301	0.0000	0.0124	0.0079		
	FAMSTAND	BRANCHE	WARBZEIT	ARBLOS	KIND	HHEINK	
o h n e	(-)	4.82	345.79	383.97	0.12	23.94	12.88
	(AGBS)	29.31	273.40	306.70	9.69	0.60	58.17
	(4 x 3)	24.65	226.37	307.65	6.11	0.25	47.48
	(6 x 2)	26.56	228.34	298.08	10.49	0.37	48.03
	(AG+SB)	25.95	234.47	302.92	12.62	0.02	39.19
	(AG)	23.30	377.91	394.63	1.15	0.02	23.39
	(BS)	20.13	247.05	329.88	4.34	17.51	34.23
Transfor- mation	(A+G+B+S)	27.80	242.79	309.08	11.10	0.00	38.53
m i t	(-)	119.51	292.51	354.66	2.48	38.32	59.22
	(AGBS)	27.54	285.68	316.99	18.77	0.04	19.17
	(4 x 3)	27.94	234.88	317.61	12.70	0.25	13.35
	(6 x 2)	17.81	235.53	293.00	18.53	0.22	10.25
	(AG+SB)	25.56	241.39	304.14	26.74	0.96	26.61
	(AG)	21.30	340.33	413.95	7.13	1.47	21.06
	(BS)	90.75	256.71	317.63	7.44	26.03	44.38
Transfor- mation	(A+G+B+S)	34.17	248.38	313.64	27.76	1.26	27.91

Anhang: IPF.ARC : Archiv-Datei als Library in GAUSS 2.0.

Die Datei enthält folgende Prozeduren:

ROT: zur Umordnung antilexikografisch angeordneter Zellhäufigkeitsdaten

MEANK: zur stückweisen Summenbildung in Spaltenvektoren

MEANKOMP: zur Summenbildung in Randtafeln

RANDVERT: zur Berechnung von Randtafeln

SDI: zur Berechnung des Soll-durch-Ist

IPF: Zur Durchführung des IPF-Algorithmus

AUSPR: zur Anordnung der Ausprägungen einer Variablen

```
proc (2)=rot(varkats, cellfreq,k);
```

```
/*-----
```

```
Format:      {vkneu,cfneu}=rot(varkats,cellfreq,k);
```

```
Purpose:      reorders antilexicographically ordered cell frequencies
```

```
Input:      varkats --> kx1 vector, where k is the number of variables,  
                  the komponents of varkat contain the number  
                  of values for each variable
```

```
            cellfreqs --> Nx1 vector, where N=prodc(varkats), containing  
                  the cell frequencies of the table ordered in  
                  antilexicographical order, i.e. the first variable  
                  chages most, the last least.
```

```
            k --->    positive integer, if k>rows(varkats), k will be  
                  replaced by k-rows(varkats).
```

```
Output:      cfneu --> Nx1 vector re-ordered cell frequencies, now with the  
                  k-th variable changing most, then the (k+1)st, etc.,  
                  the (k-1)st slowest (mod rows(varkats)).
```

```
            vkneu ---> Nx1 vector,  varkats rotated suitable to cfneu.
```

```
-----*/
```

```
local v,r1,c1,b,c;
```

```
if k>rows(varkats);
```

```
    k=k-rows(varkats);
```

```
endif;
```

```
if k==1;
```

```
    retp(varkats,cellfreq);
```

```
else;
```

```
    v=cumprodc(varkats);
```

```
    c1=v[k-1];
```

```
    r1=rows(cellfreq)/c1;
```

```
    b=vec(reshape(cellfreq,r1,c1));
```

```
    c=rotater(varkats',-(k-1))';
```

```
    retp(c,b);
```

```
endif;
```

```
endp;
```

```

proc (1)=meank(k,vektor);
/*-----*/
Format:      vneu = meank(k,vektor)

Purpose:     calculates means of k successive elements of vektor

Input:       k --->  positive integer

              vektor ---> nx1 vektor, n being a multiple of k

Output:      vneu --->  nx1 vektor, the (mk+i)-th element of vektor
                      replaced by the mean of
                      vektor[mk+1],...,vektor[mk+k].

/*-----*/

local a;
a=reshape(vektor,rows(vektor)/k,k);
a=meanc(a') .* ones(1,k);
retp(vec(a'));
endp;

proc (1)=meankomp(varkats,cellfreq,k);
/*-----*/
Format:      meancell = meankomp(varkats,cellfreq,k);

prupose:     calculates means within a table

Input:       varkats, cellfreq, k ---> as in ROT

Output:      meancell --->replaces each cell frequency by its mean frequency
                      with respect to the k-th variable

/*-----*/

local v,c;
{v,c}=rot(varkats,cellfreq,k);
c=meank(v[1],c);
{v,c}=rot(v,c,rows(varkats)+2-k);
retp(c);
endp;

```

```

proc (1)=randvert(varkats,cellfreq,randvec);
/*-----
Format:      mcell=randvert(varkats,cellfreq,randvec);

Purpose:      calculates boundary tables

Input:        varkats
              cellfreq ---> as in ROT

              randvec ---> kx1 vektor (k=rows(varkats)) of 0's and 1's.

Output:       mcell ---> Nx1 vector, where the cell frequencies of
                      cellfreq are replaced by the means with
                      respect to variables indicated by 1 in
                      randvec.

-----*/

local i,c;
i=1;c=cellfreq;
do until i>rows(randvec);
if randvec[i]==1;
    c=meankomp(varkats,c,i);
endif;
i=i+1;
endo;
retp(c);
endp;

proc (1)=sdi(soll,ist);
/*-----
Format:      quotient=sdi(soll,ist)
Purpose:      Calculates "Soll durch Ist"
Input:        "soll"
              "ist" ---> Nx1-vectors
Output:       ratio soll./ist , if ist./=0
              soll    otherwise

-----*/

local quot;
quot=(soll ./ (ist + (ist==0)));
retp(quot);
endp;

```



```

proc (2) = ipf(varkats,design,startvec,sollvec,eps);
/*-----
Format:      {fitdata,schleife} = ipf(varkats,design,startvec,sollvec,eps);
Purpose:     Iterative proportional fitting
Input:       varkats --> kx1 vector, where k is the number of variables,
                    the komponents of varkat contain the number
                    of values for each variable
            design ---> kxm matrix of 0's and 1's. Each column denotes
                    a boundary table to be fitted; a 1 for each
                    variable to be contains in this boundary table.
            startvec --> cell frequencies, used as starting table
            sollvec ---> cell frequencies, wohse boundary distributions
                    according to design shall be fitted
            eps ---> tolerance.
                    The iterative procedure stops, if the maximum
                    distance between two fits is less than
                    eps*maxc(startvec)

Output:      schleife ---> number of loops until iterations stopped
            fitdata ---> final cell frequencies after iteration
-----*/
local distanz,i,fitdata,fitalt,schleife;
if rows(varkats) /= rows(design); "Fehler in Proc IPF: Zeile1"; wait;endif;
if rows(startvec) /= rows(sollvec); "Fehler in Proc IPF: Zeile
2";wait;endif;
if rows(startvec) /= prodc(varkats); "Fehler in Proc IPF: Zeile 3";
wait;endif;
cls;"Iterative proportional fitting procedure";
"Design:";
design;
distanz=sumc(startvec);
sollvec=sollvec.*(sumc(startvec)/sumc(sollvec));
fitdata=startvec;
fitalt=fitdata;
schleife=0;
do until distanz<eps*maxc(startvec);
    i=1;
    do until i>cols(design);
        fitdata=sdi(randvert(varkats,sollvec,1-design[.,i]),
                    randvert(varkats,fitdata,1-design[.,i]))
                    .* fitdata;

        i=i+1;
    endo;
    fitdata=fitdata .* (sumc(startvec) / sumc(fitdata));
    distanz= maxc(abs(fitdata-fitalt));
    schleife=schleife+1;
    fitalt=fitdata;
    i=1;
    "Schleife ";;schleife;;": Distanz=";;distanz;;chrs(13));
end;
cls;
"Ende des IPF nach ";; schleife;; " Schleifen.";
retp(schleife,fitdata);
endp;

```

```

proc (1)=auspr(x);
/*-----
Format      u=auspr(x);

Input:      x ---->  nx1- vektor

Output:     u ---->  kx1 vector, containing the different values of x in
                    increasing order.
-----*/
/*local y,z;
y=sortc(x,1);
z=y[2:rows(y),.]-y[1:(rows(y)-1),.];
retp(y[1,.]!delif(y[2:rows(y),.],z.==0));
endp;
*/
local z;
z=x[1];
do until x==x[1];
    x=delif(x,x.==x[1]);
    z=z!x[1];
enddo;
z=sortc(z,1);
retp(z);
endp;

```


ZUMA-Arbeitsberichte

- 80/15 Gerhard Arminger, Willibald Nagl, Karl F. Schuessler
Methoden der Analyse zeitbezogener Daten. Vortragsskripten der ZUMA-
Arbeitstagung vom 25.09. - 05.10.79
- 81/07 Erika Brückner, Hans-Peter Kirschner, Rolf Porst, Peter Prüfer, Peter
Schmidt
Methodenbericht zum "ALLBUS 1980"
- 81/19 Manfred Küchler, Thomas P. Wilson, Don H. Zimmerman
Integration von qualitativen und quantitativen Forschungsansätzen
- 82/03 Gerhard Arminger, Horst Busse, Manfred Küchler
Verallgemeinerte Lineare Modelle in der empirischen Sozialforschung
- 82/08 Glenn R. Carroll
Dynamic analysis of discrete dependent variables: A didactic essay
- 82/09 Manfred Küchler
Zur Messung der Stabilität von Wählerpotentialen
- 82/10 Manfred Küchler
Zur Konstanz der Recallfrage
- 82/12 Rolf Porst
"ALLBUS 1982" - Systematische Variablenübersicht und erste Ansätze zu
einer Kritik des Fragenprogramms
- 82/13 Peter Ph. Mohler
SAR - Simple AND Retrieval mit dem Siemens-EDT-Textmanipulationspro-
gramm
- 82/14 Cornelia Krauth
Vergleichsstudien zum "ALLBUS 1980"
- 82/21 Werner Hagstotz, Hans-Peter Kirschner, Rolf Porst, Peter Prüfer
Methodenbericht zum "ALLBUS 1982"
- 83/09 Bernd Wegener
Two approaches to the analysis of judgments of prestige: Interindivi-
dual differences and the general scale
- 83/11 Rolf Porst
Synopsis der ALLBUS-Variablen. Die Systematik des ALLBUS-Fragenpro-
gramms und ihre inhaltliche Ausgestaltung im ALLBUS 1980 und ALLBUS
1982
- 84/01 Manfred Küchler, Peter Ph. Mohler
Qualshop (ZUMA-Arbeitstagung zum "Datenmanagement bei qualitativen
Erhebungsverfahren") - Sammlung von Arbeitspapieren und -berichten,
Teil I + II
- 84/02 Bernd Wegener
Gibt es Sozialprestige? Konstruktion und Validität der Magnitude-
Prestige-Skala

- 84/03 Peter Prüfer, Margrit Rexroth
Erfahrungen mit einer Technik zur Bewertung von Interviewerverhalten
- 84/04 Frank Faulbaum
Ergebnisse der Methodenstudie zur internationalen Vergleichbarkeit von Einstellungsskalen in der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) 1982
- 84/05 Jürgen Hoffmeyer-Zlotnik
Wohnquartiersbeschreibung. Ein Instrument zur Bestimmung des sozialen Status von Zielhaushalten
- 84/07 Gabriele Hippler, Hans-Jürgen Hippler
Reducing Refusal Rates in the Case of Threatening Questions: The "Door-in-the-Face" Technique
- 85/01 Hartmut Esser
Befragtenverhalten als "rationales Handeln" - Zur Erklärung von Antwortverzerrungen in Interviews
- 85/03 Rolf Porst, Peter Prüfer, Michael Wiedenbeck, Klaus Zeifang
Methodenbericht zum "ALLBUS 1984"
- 86/01 Dagmar Krebs
Zur Konstruktion von Einstellungsskalen im interkulturellen Vergleich
- 86/02 Hartmut Esser
Können Befragte lügen? Zum Konzept des "wahren Wertes" im Rahmen der handlungstheoretischen Erklärung von Situationseinflüssen bei der Befragung
- 86/03 Bernd Wegener
Prestige and Status as Function of Unit Size
- 86/04 Frank Faulbaum
Very Soft Modeling: The Logical Specification and Analysis of Complex Process Explanations with Arbitrary Degrees of Underidentification and Variables of Arbitrary Aggregation and Measurement Levels
- 86/05 Peter Prüfer, Margrit Rexroth (Übersetzung: Dorothy Duncan)
On the Use of the Interaction Coding Technique
- 86/06 Hans-Peter Kirschner
Zur Kessler-Greenberg-Zerlegung der Varianz der Meßdifferenz zwischen zwei Meßzeitpunkten einer Panel-Befragung
- 86/07 Georg Erdmann
Ansätze zur Abbildung sozialer Systeme mittels nicht-linearer dynamischer Modelle
- 86/09 Heiner Ritter
Einige Ergebnisse von Vergleichstests zwischen den PC- und Mainframe-Versionen von SPSS und SAS
- 86/11 Günter Rothe
Bootstrap in generalisierten linearen Modellen
- 87/01 Klaus Zeifang
Die Test-Retest-Studie zum ALLBUS 1984 - Tabellenband

- 87/02 Klaus Zeifang
Die Test-Retest-Studie zum ALLBUS 1984 - Abschlußbericht
- 87/04 Barbara Erbslöh, Michael Wiedenbeck
Methodenbericht zum "ALLBUS 1986"
- 87/05 Norbert Schwarz, Julia Bienias
What Mediates the Impact of Response Alternatives on Behavioral Reports?
- 87/06 Norbert Schwarz, Fritz Strack, Gesine Müller, Brigitte Chassein
The Range of Response Alternatives May Determine the Meaning of the Question: Further Evidence on Informative Functions of Response Alternatives
- 87/07 Fritz Strack, Leonard L. Martin, Norbert Schwarz
The Context Paradox in Attitude Surveys: Assimilation or Contrast?
- 87/08 Gudmund R. Iversen
Introduction to Contextual Analysis
- 87/09 Seymour Sudman, Norbert Schwarz
Contributions of Cognitive Psychology to Data Collection in Marketing Research
- 87/10 Norbert Schwarz, Fritz Strack, Denis Hilton, Gabi Naderer
Base-Rates, Representativeness, and the Logic of Conversation
- 87/11 George F. Bishop, Hans-Jürgen Hippler, Norbert Schwarz, Fritz Strack
A Comparison of Response Effects in Self-Administered and Telephone Surveys
- 87/12 Norbert Schwarz
Stimmung als Information. Zum Einfluß von Stimmungen und Emotionen auf evaluative Urteile
- 88/01 Antje Nebel, Fritz Strack, Norbert Schwarz
Tests als Treatment: Wie die psychologische Messung ihren Gegenstand verändert
- 88/02 Gerd Böhner, Herbert Bless, Norbert Schwarz, Fritz Strack
What Triggers Causal Attributions? The Impact of Valence and Subjective Probability
- 88/03 Norbert Schwarz, Fritz Strack
The Survey Interview and the Logic of Conversation: Implications for Questionnaire Construction
- 88/04 Hans-Jürgen Hippler, Norbert Schwarz
"No Opinion"-Filters: A Cognitive Perspective
- 88/05 Norbert Schwarz, Fritz Strack
Evaluating One's Life: A Judgment of Subjective Well-Being
- 88/06 Norbert Schwarz, Herbert Bless, Gerd Böhner, Uwe Harlacher, Margit Kellenbenz
Response Scales as Frames of Reference:
The Impact of Frequency Range on Diagnostic Judgments

- 88/07 Michael Braun
Allbus-Bibliographie (7. Fassung, Stand: 30.6.88)
- 88/08 Günter Rothe
Ein Ansatz zur Konstruktion inferenzstatistisch verwertbarer Indices
- 88/09 Ute Hauck, Reiner Trometer
Methodenbericht
International Social Survey Program - ISSP 1987
- 88/10 Norbert Schwarz
Assessing frequency reports of mundane behaviors:
Contributions of cognitive psychology to questionnaire
construction
- 88/11 Norbert Schwarz, B. Scheuring (sub.)
Judgments of relationship satisfaction: Inter- and intraindividual
comparison strategies as a function of questionnaire structure
- 88/12 Rolf Porst, Michael Schneid
Ausfälle und Verweigerungen bei Panelbefragungen
- Ein Beispiel -
- 88/13 Cornelia Züll
SPSS-X. Anmerkungen zur Siemens BS2000 Version
- 88/14 Michael Schneid
Datenerhebung am PC - Vergleich der Interviewprogramme "interv⁺"
und "THIS"
- 88/15 Norbert Schwarz, Bettina Scheuring
Die Vergleichsrichtung bestimmt das Ergebnis
von Vergleichsprozessen:
Ist - Idealdiskrepanzen in der Partnerwahrnehmung
- 89/01 Norbert Schwarz, George F. Bishop, Hans-J. Hippler, Fritz Strack
Psychological Sources Of Response Effects in Self-Administered
And Telephone Surveys
- 89/02 Michael Braun, Reiner Trometer, Michael Wiedenbeck,
Methodenbericht. Allgemeine Bevölkerungsumfrage der
Sozialwissenschaften - ALLBUS 1988 -
- 89/03 Norbert Schwarz
Feelings as Information:
Informational and Motivational Functions of Affective States
- 89/04 Günter Rothe
Jackknife and Bootstrap:
Resampling-Verfahren zur Genauigkeitsschätzung
von Parameterschätzungen
- 89/05 Herbert Bless, Gerd Bohner, Norbert Schwarz und Fritz Strack
Happy and Mindless?
Moods and the Processing of Persuasive Communications
- 89/06 Gerd Bohner, Norbert Schwarz und Stefan E. Hornuth
Die Stimmungs-Skala: Eine deutsche Version des "Mood Survey"
von Underwood und Froming

- 89/07 Ulrich Mueller
Evolutionary Fundamentals of Social Inequality, Dominance
and Cooperation
- 89/08 Robert Huckfeldt
Noncompliance and the Limits of Coercion:
The Problematic Enforcement of Unpopular Laws
- 89/09 Peter Ph. Mohler, Katja Frehsen und Ute Hauck
CUI - Computerunterstützte Inhaltsanalyse
Grundzüge und Auswahlbibliographie zu neueren Anwendungen
- 89/10 Cornelia Züll, Peter Ph. Mohler
Der General Inquirer III -
Ein Dinosaurier für die historische Forschung
- 89/11 Fritz Strack, Norbert Schwarz, Brigitte Chassein, Dieter Kern,
Dirk Wagner
The Salience of Comparison Standards and the Activation of
Social Norms: Consequences for Judgments of Happiness and their
Communication
- 89/12 Jutta Kreiselmaier, Rolf Porst
Methodische Probleme bei der Durchführung telefonischer
Befragungen: Stichprobenziehung und Ermittlung von Zielpersonen,
Ausschöpfung und Nonresponse, Qualität der Daten.
- 89/13 Rainer Mathes
Modulsystem und Netzwerktechnik.
Neuere inhaltsanalytische Verfahren zur Analyse von
Kommunikationsinhalten.
- 89/14 Jutta Kreiselmaier, Peter Prüfer, Margrit Rexroth
Der Interviewer im Pretest.
Evaluation der Interviewerleistung und Entwurf eines
neuen Pretestkonzepts. April 1989.
- 89/15 Henrik Tham
Crime as a Social Indicator.
- 89/16 Ulrich Mueller
Expanding the Theoretical and Methodological Framework of
Social Dilemma Research
- 89/17 Hans-J. Hippler, Norbert Schwarz, Elisabeth Noelle-Neumann
Response Order Effects in Dichotomous Questions:
The Impact of Administration Mode
- 89/18 Norbert Schwarz, Hans-J. Hippler, Elisabeth Noelle-Neumann,
Thomas Münkler
Response Order Effects in Long Lists:
Primacy, Recency, and Asymmetric Contrast Effects
- 89/19 Wolfgang Meyer
Umweltberichterstattung in der Bundesrepublik Deutschland
- 89/20 Michael Braun, Reiner Trometer
ALLBUS Bibliographie (8. Fassung, Stand: 30.6. 1989)